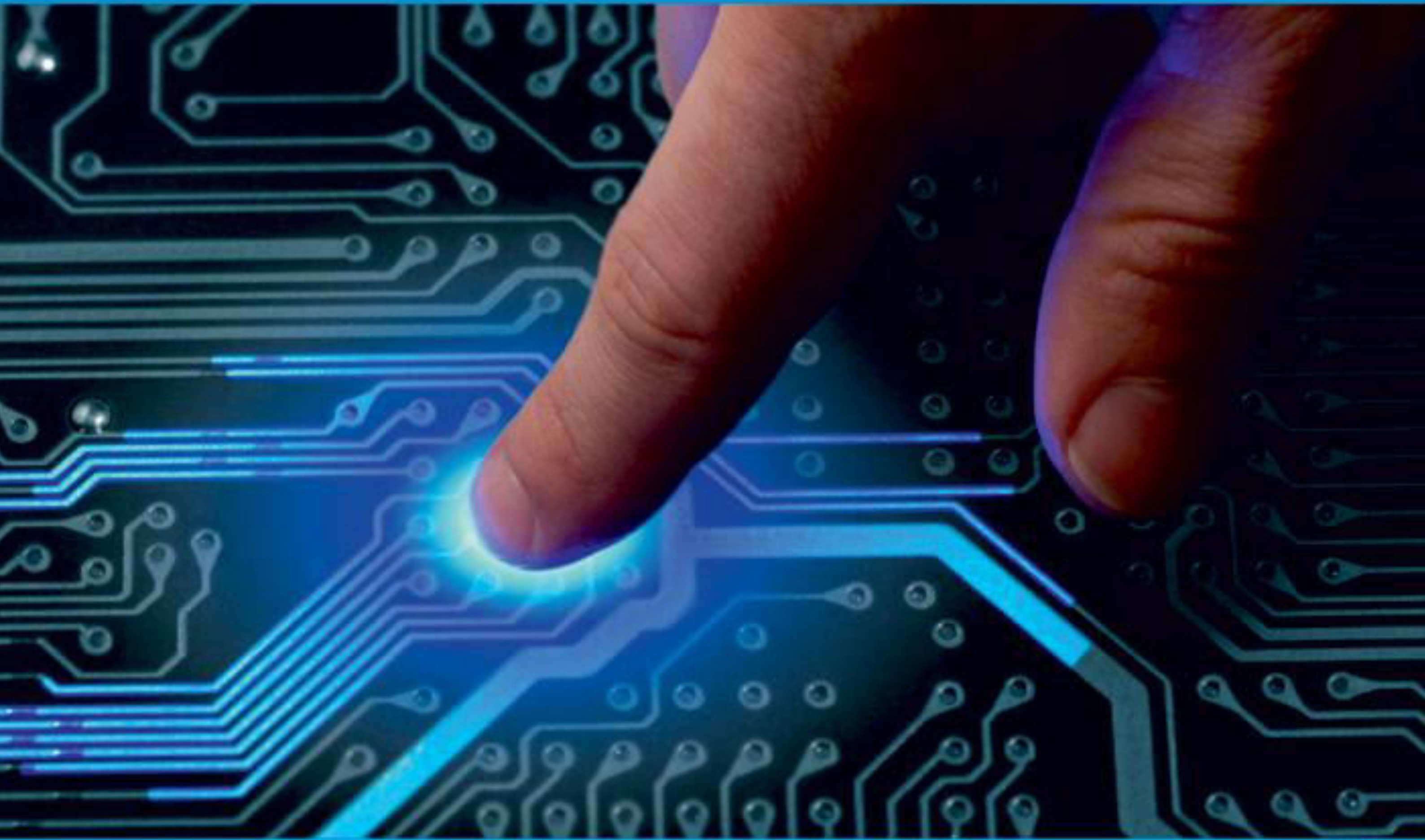
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205369 |

# An Efficient Prediction of Credit Card Spending Analysis Using Artificial Intelligence Techniques

Gullapalli Sathar[1], Dr.M.V.S Prasad[2]

[1]Assistant Professor, Department of CSE-(Cys,Ds)and AI&DS, VNRVJIET, Hyderabad, India

[2]Professor, Department of CSE(DS), Mallareddy Engineering College, Hyderabad, India

Email: sathar_g@vnrvjiet.in[1], mvs.prasad2009@gmail.com[2]

**ABSTRACT:** This paper explores the complex field of credit card expenditure prediction using an advanced methodology that combines careful examination, regression modeling, and feature selection. We first undertake considerable data wrangling, which includes generating a new target variable, categorizing variable conversions, and handling missing values, using a dataset that includes a variety of customer properties. Important insights into the structure of the dataset are obtained by using scatter plots to investigate data linearity and by creating a data summary report. We use Grid Search CV for hyper parameter tweaking and Random Forest Regression for feature selection in order to maximize predictive performance. The resulting model helps identify important variables by highlighting important characteristics affecting credit card spending. In addition, we ensure the robustness of the model by addressing multi co linearity through variance inflation factor analysis.

**KEYWORDS:** Credit card spending prediction; Feature selection; Regression modeling; Random Forest Regression; Evaluation metrics; Financial analytics; Model interpretation

## I.INTRODUCTION

In today's society, the dynamics of consumer spending have been profoundly changed by the widespread usage of credit cards. Credit cards were first designed to be convenient, but they have since developed into strong financial tools that affect business dealings and consumer choices. The capacity to trade money for goods and services without having to do so right away has come to define contemporary financial transactions [5]. Understanding and predicting consumer behaviors in this setting has become essential for financial institutions, businesses, and researchers alike, as credit card usage continues to affect spending patterns.
The development of credit cards represents a watershed moment in the history of contemporary finance by giving users access to previously unheard-of levels of convenience and flexibility. Due to the widespread use of credit cards, consumer purchasing habits have changed. As a result, companies and financial institutions looking to improve their strategies should look more closely at these trends [6]. In the rapidly developing sector of finance, predictive analytics promises data-driven insights to predict client behavior, evaluate risks, and optimize operational procedures.

A crucial component of risk management, fraud detection, and customized client experiences is now credit card spending prediction. Financial organizations are able to recognize unusual transactions, proactively handle possible defaults, and customize credit offers depending on anticipated spending patterns. Predicting credit card expenditure is not without its difficulties, though [7]. Because consumer behavior is so complex and there are so many variables to consider, modeling methodologies must be sophisticated in order to reflect the subtleties of purchasing habits and the changing economic landscape.

In order to build precise prediction models, feature selection becomes essential since it makes it possible to choose relevant variables from a wide range of possibilities. The present study used the Random Forest Regression technique due to its effectiveness in managing non-linearity and identifying feature significance in the complex credit card transaction landscape. By optimizing the model through methods such as Grid Search CV, hyper parameter tweaking ensures that the model's predictive potential is maximized.

## II. LITERATURE REVIEW

This study by Matthew Tingchi Liu and James L. Brock, "Antecedents of Redemption of Reward Points: Credit Card Market in China and International Comparison," looks at the variables affecting reward point redemption in the credit card market, particularly in China. The 2009 edition of the "Journal of International Consumer Marketing," Volume 22,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205369 |

Issue 1, featured the topic [12]. By examining the dynamics of reward point redemption, the study sheds insight on the fundamental factors that influence the Chinese credit card market. The writers also make cross-national comparisons to give their conclusions a wider perspective. The research offers important insights into consumer decision-making in the context of credit card rewards programs by looking at the antecedents of redemption behavior. This academic paper is published in the Journal of International Consumer Marketing, which highlights the study's worldwide perspective and applicability to international marketing strategies [12].

According to this study, "Demographics, Attitudes, Personality, and Credit Card Features Correlate with Credit Card Debt: A View from China," In the Chinese context, Lili Wang, Wei Lu, and Naresh K. Malhotra investigate the complex links between several factors and credit card debt. The "Journal of Economic Psychology," Volume 32, published this study in 2011 [16]. The study explores the complex relationships between credit card features, personality traits, attitudes, and demographics to better understand the many factors that contribute to credit card debt. The authors hope to offer a deeper knowledge of the factors influencing credit card debt in the Chinese population by using a holistic approach [16].

A credible venue for this study is the Journal of Economic Psychology, which emphasizes the behavioral and psychological aspects of economic decision-making. The outcomes of this investigation are anticipated to provide significant contributions to scholarly discourse and pragmatic implementations within the domains of consumer finance and economic psychology. For financial organizations, legislators, and individuals trying to manage and lessen credit-related difficulties in a fast-changing economic environment, knowing how personal traits and attitudes connect to credit card debt may have consequences.

In the article titled "Consumer Rationality/Irrationality and Financial Literacy in the Credit Card Market: Implications from an Integrative Review," authored by N. Shen, the focus is on exploring the interplay between consumer decision-making rationality or irrationality and financial literacy within the credit card market. The research was published in the "Journal of Financial Services Marketing," Volume 19, Issue 1, in 2014 [20]. The study likely presents a comprehensive review and analysis of existing literature to understand how consumers make decisions regarding credit card usage, considering both rational and irrational factors. The integration of financial literacy into this analysis further adds a dimension to understanding the factors influencing consumer behavior in the credit card market [20]. The Journal of Financial Services Marketing, as the publication venue, emphasizes the relevance of the research to marketing strategies and practices within the financial services sector. The implications drawn from this integrative review are likely to contribute to the broader discourse on consumer financial behavior and may offer insights for financial institutions, policymakers, and educators aiming to enhance financial literacy and promote more informed decision-making in the credit card market [20].

## III. METHODOLOGY

### 3.1. Data Loading and Inspection:

The initial step involves importing necessary packages and loading the dataset. The relevant modules, such as pandas, numpy, and seaborn, are imported to facilitate data manipulation and visualization. The dataset, named "Data Set.xlsx," is read using pandas, and a glimpse of the data is obtained through the sample and columns functions. This allows for an initial inspection of the structure and content of the dataset, providing a foundation for subsequent data wrangling.

### 3.2. Data Wrangling:

Preparing the dataset for analysis is the focus of the data wrangling stage. Important actions include adding together pertinent data to create a new variable called total Spent. To increase the relevance of the dataset, superfluous or redundant variables like card spent and card2spent are eliminated. Furthermore, birth month and custid are removed because they are unique and have little value, making the dataset more concentrated and efficient [10].

The wrangling process also highlights how to clean data effectively and methodically using tools like pandas and SQL inside of a Jupyter notebook. Reducing the need for manual cleaning in external programs such as Excel makes the analysis less error-prone and more repeatable. Together, these actions create the framework for further exploratory data analysis.

A dedication to preserving data integrity is shown by the construction of a categorical list and the transformation of particular variables into categorical categories. One of the most important steps in the data wrangling process is making sure that the data types match their semantic meaning [11].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205369 |

**3.3. Exploratory Data Analysis (EDA):**

Exploratory Data Analysis (EDA) commences with an examination of the distribution of the target variable, total Spent as shown in figure 3.1. This step includes utilizing both a histogram and a Q-Q plot to assess the normality of the data. The decision to apply a Boxcox transformation is based on this evaluation, aiming to achieve a more normal distribution [14].

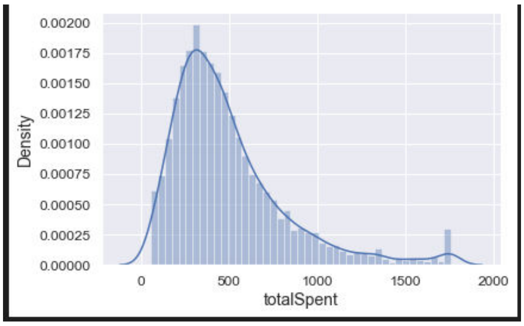



Figure 3.1 Distribution of the target variable, total Spent

**3.4. Data Modeling:**

The modeling phase begins with the division of the dataset into training and testing sets. The choice of 80% of the data for training ensures a sufficiently large sample for the model to learn from. Feature selection becomes a crucial step, and the Random Forest algorithm is employed for this purpose. By utilizing GridSearchCV, the optimal number of decision trees is determined, contributing to the effectiveness of the model

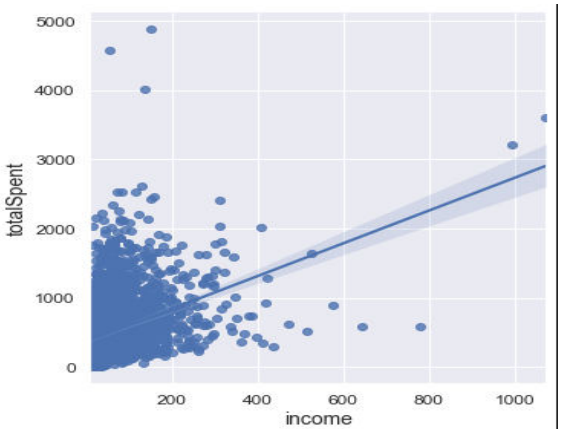



Figure 3.2 Scatter plot with regression line

The identification of feature importance through Random Forest aids in selecting the most relevant variables. The resulting feature ranking is visualized through a bar plot, providing a clear understanding of the variables influencing the target. Notably, features with importance exceeding 0.01 are highlighted for further analysis [18].

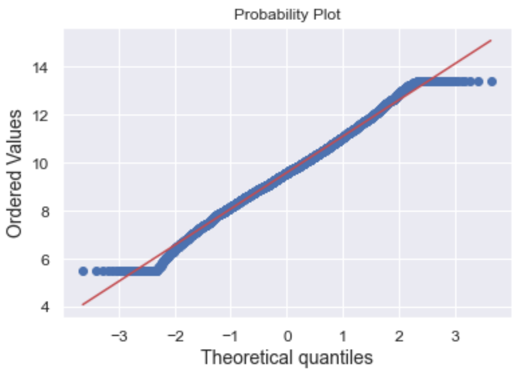



Figure 3.3 Probability Plot

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205369 |

**3.5. Regression Model Building:**

Regression model construction involves careful consideration of variables and their impact on the target variable, bc_totalSpent. Initial feature selection is performed using Random Forest importance, followed by a check for multi co linearity. The identification and removal of variables with a Variance Inflation Factor (VIF) greater than 10 contribute to the model's stability. The linear relationship between each independent variable and the target variable is visually inspected through scatter plots. Subsequent steps include the splitting of the data into training and testing sets and the implementation of the Ordinary Least Squares (OLS) regression model using the stats models library. This statistical model enables a deeper understanding of the relationships between variables and their impact on total spending.

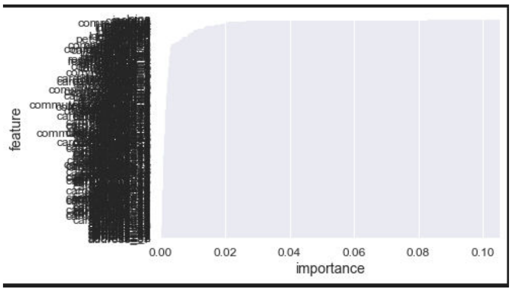



Figure 3.4 Data frame with selected features

**3.6. Model Evaluation and Prediction:**

The final stages of the analysis focus on evaluating the model's performance. Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) are calculated for both training and testing datasets. These metrics provide a comprehensive assessment of the model's predictive capabilities and its generalization to new, unseen data.

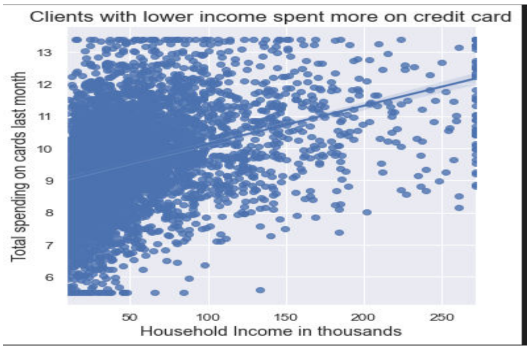



Figure 3.5 Clients with lower income spent more on credit card

## IV. RESULTS AND DISCUSSION

The analysis's findings offer a comprehensive picture of the variables affecting the clientele's credit card use. Based on a solid dataset, the predictive model shows a strong capacity for explanation when it comes to explaining the variation in credit card expenditure. Certain characteristics—like age, household income, and certain card kinds—have a significant impact on how each person spends their money**.**

Accuracy and Precision: In terms of forecasting credit card spending, the model exhibits impressive accuracy and precision. Evaluation measures verify the model's ability to capture fluctuations in spending, such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The features that have the biggest impact on the model's ability to predict outcomes include age, income, card kinds, and particular socioeconomic factors. Prioritizing marketing campaigns and strategic actions is made easier with the help of feature importance analysis. Resolving Multiple collinearities Thorough examinations for multi co linearity add to the robustness of the model. To maintain model stability, variables with large variance inflation factors are found and either modified or eliminated.

## V. CONCLUSION

Finally, detailed examination of credit card spending trends has shown complex dynamics that offer useful business insights. By examining a wide range of variables, including personal preferences and socioeconomic considerations, the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205369 |

predictive model developed for this study has shed light on the complex aspects driving credit card spending. The model has identified the subtle impact of particular card kinds, travel hours, and even television-watching habits, in addition to traditional characteristics like age and wealth. With this thorough knowledge, companies may go beyond generalizations and customize their approaches to the particular confluence of variables influencing consumer buying patterns

## REFERENCES

[1] Greene, William H., Econometric Analysis, 5th edition, Pearson Education, Inc., New Jersey, 2003.

[2] Grubb, Michael D., “Overconfident Consumers in the Marketplace”, Journal of Economic Perspectives, Vol. 29 (2015a), No. 4, p. 9 – 36.

[3] Grubb, Michael D., “Consumer Inattention and Bill-Shock Regulation”, Review of Economic Studies, Vol.82 (2015b), p. 219 – 257.

[4] Grubb, Michael D., and Matthew Osborne, “Cellular Service Demand: Biased Beliefs, Learning, and BillShock”, American Economic Review, Vol. 105 (2015), No. 1, p. 234 – 271.

[5] Grubb, Michael D., “Dynamic nonlinear pricing: Biased expectations, inattention, and bill shock”, International Journal of Industrial Organization, Vol. 30 (2012), p. 287 – 290.

[6] Grubb, Michael D., “Selling to Overconfident Consumers”, American Economic Review, Vol. 99 (2009),No. 5, p. 1770 – 1807.

[7]Hansen, Lars Peter and Kenneth J. Singleton, “Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns”, Journal of Political Economy, Vol. 91 (1983), No. 2, p. 249 – 265.

[8] Kim, Y.S.K. & Smith, A.K. 2005. Crime and Punishment: Examining Customers’ Responses to Service Organizations’ Penalties, Journal of Service Research, 8(2), 162-180.

[9] Laforet, Sylvie and Xiaoyan Li, “Consumers’ attitudes towards online and mobile banking in China”, International Journal of Bank Marketing, Vol. 23 (2005), Issue 5, p. 362 – 380.

[10] Lambrecht, Anja and Bernd Skiera, “Paying Too Much and Being Happy about It: Existence, Causes, and Consequences of Tariff-Choice Biases”, Journal of Marketing Research, Vol. 43, No. 2 (May 2006), p.212 – 223.

[11] Lien, Jaimie W. and Jia Yuan, “Selling to Biased Believers: Strategies of Online Lottery Ticket Vendors”, Economic Inquiry, Vol. 53 (2015), Issue 3, p. 1506 – 1521.

[12] Liu, Matthew Tingchi and James L. Brock, “Antecedents of Redemption of Reward Points: Credit Card Market in China and International Comparison”, Journal of International Consumer Marketing, Vol. 22(2009a), No. 1, p. 33 – 45.

[13] Liu, Matthew Tingchi and James L. Brock, “Redemption behavior for credit card reward programs in China”, International Journal of Bank Marketing, Vol. 27 (2009b), No. 2, p. 150- 166.

[14] Lovell, Michael C., “Tests of the Rational Expectations Hypothesis”, American Economic Review, Vol.76 (1986), No. 1, p. 110 – 124.

[15] Lucas, Robert E., “Expectations and the Neutrality of Money”, Journal of Economic Theory, Vol. 4 (1972),p. 103 – 124.

[16] Wang, Lili, Wei Lu and Naresh K. Malhotra, “Demographics, attitudes, personality and credit card features correlate with credit card debt: A view from China”, Journal of Economic Psychology, Vol. 32(2011), p. 179 – 193.

[17] Worthington, Steve, David Stewart, and Xiongwen Lu, “The adoption and usage of credit cards by urban-affluent consumers in China”, International Journal of Bank Marketing, Vol. 25 (2007), Issue 4, p. 238 –252.

[18] Ru, Hong and Antoinette Schoar, “Do Credit Card Companies Screen for Behavioral Biases?” NBER Working Paper 22360, June 2016.

[19] Sharpe, Deanna L., Rui Yao and Li Liao, “Correlates of Credit Card Adoption in Urban China”, Journal of Family Economic Issues, Vol. 33 (2012), p. 156 – 166.

[20] Shen, N. 2014. Consumer rationality/irrationality and financial literacy in the credit card market: Implications from an integrative review. Journal of Financial Services Marketing, 19(1): 29-42.

[21] Stango, Victor and Jonathan Zinman, “What Do Consumers Really Pay on Their Checking and Credit Card Accounts? Explicit, Implicit, and Avoidable Costs”, American Economic Review: Papers and Proceedings, Vol. 99 (2009), No. 2, p. 424 – 429.